



OPEN ACCESS

EDITED BY

Ruzong Fan,
Georgetown University Medical Center,
United States

REVIEWED BY

Cen Wu,
Kansas State University, United States
Chi Zhang,
Yale University, United States

*CORRESPONDENCE

Wan-Yu Lin,
✉ linwy@ntu.edu.tw

RECEIVED 24 April 2025

ACCEPTED 04 August 2025

PUBLISHED 01 September 2025

CITATION

Lin W-Y (2025) Mining for gene-environment and gene-gene interactions: parametric and non-parametric tests for detecting variance quantitative trait loci.
Front. Genet. 16:1617504.
doi: 10.3389/fgene.2025.1617504

COPYRIGHT

© 2025 Lin. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mining for gene-environment and gene-gene interactions: parametric and non-parametric tests for detecting variance quantitative trait loci

Wan-Yu Lin 1,2*

¹Institute of Health Data Analytics and Statistics, College of Public Health, National Taiwan University, Taipei, Taiwan, ²Master of Public Health Program, College of Public Health, National Taiwan University, Taipei, Taiwan

Introduction: Detection of variance quantitative trait loci (vQTL) can facilitate the discovery of gene-environment (GxE) and gene-gene interactions (GxG). Identifying vQTLs before direct GxE and GxG analyses can considerably reduce the number of tests and the multiple-testing penalty.

Methods: Despite some methods proposed for vQTL detection, few studies have performed a head-to-head comparison simultaneously concerning false positive rates (FPRs), power, and computational time. This work compares three parametric and two non-parametric vQTL tests.

Results: Simulation studies show that the deviation regression model (DRM) and Kruskal-Wallis test (KW) are the most recommended parametric and non-parametric tests, respectively. The quantile integral linear model (QUAIL, non-parametric) appropriately preserves the FPR under normally or non-normally distributed traits. However, its power is never among the optimal choices, and its computational time is much longer than that of competitors. The Brown-Forsythe test (BF, parametric) can suffer from severe inflation in FPR when SNP's minor allele frequencies <0.2. The double generalized linear model (DGLM, parametric) is not valid for non-normally distributed traits, although it is the most powerful method for normally distributed traits.

Discussion: Considering the robustness (to outliers) and computation time, I chose KW to analyze four lipid traits in the Taiwan Biobank. I further showed that GxE and GxG were enriched among 30 vQTLs identified from the four lipid traits.

KEYWORDS

cholesterol, dyslipidemia, triglyceride, variance quantitative trait locus, Taiwan Biobank

1 Introduction

Omitting essential predictors from a regression model can lead to heteroscedasticity (Breusch and Pagan, 1979). In genetic analyses, loci with unequal phenotypic variances across different genotype groups are called “variance quantitative trait loci” (vQTLs), which can be caused by omitting gene-environment interaction (GxE) or gene-gene interaction (GxG) in the regression model (Pare et al., 2010; Ronnegard and Valdar, 2012; Shi, 2022; Wang et al., 2019). For example, Wang et al. performed a genome-wide vQTL analysis of 5.6 million variants on ~350,000 unrelated individuals of European ancestry for 13 quantitative traits. They identified 75 significant vQTLs for nine traits, especially

those related to obesity. Pervasive GxE effects for obesity-related traits were further explored through direct GxE analyses. Wang et al.'s study has demonstrated the detection of GxE without environmental data (Wang et al., 2019).

Methods for detecting vQTLs are categorized as parametric or non-parametric. The development of parametric vQTL tests usually depends on the assumption of normally distributed traits (Marderstein et al., 2021; Young et al., 2018). Contrastingly, non-parametric vQTL tests are generally more robust to trait distributions (Miao et al., 2022).

For every single-nucleotide polymorphism (SNP), we can use the Brown-Forsythe (BF) test (Brown and Forsythe, 1974) to compute the dispersion within three genotype groups. Compared with the Levene's test (Levene, 1960), the BF test is more robust to outliers by choosing the median to replace the mean as the center of each genotype group. Moreover, a "deviation regression model" (DRM) has been proposed to allow continuous predictors such as minor allele dosages (Marderstein et al., 2021). Therefore, the predictor is not limited to minor allele counts of SNPs (i.e., 0, 1, and 2). Furthermore, a double generalized linear model (DGLM) was also developed to identify loci associated with trait variability and to detect interactions in genome-wide association studies (GWAS) (Ronnegard and Valdar, 2011; Smyth, 1989). BF, DRM, and DGLM are all the so-called parametric vQTL methods.

Non-parametric methods for vQTL identification include the Kruskal-Wallis test (KW) (Kruskal and Wallis, 1952) and a quantile integral linear model (QUAIL) (Miao et al., 2022). As the non-parametric counterpart of one-way analysis of variance (ANOVA), KW is used to compare the median difference between multiple groups. As long as we first calculate the deviation between trait values and within-group medians, KW can also be applied to test phenotypic homoscedasticity among the three genotype groups. Another non-parametric method, QUAIL, assesses genetic effects on phenotypic variability based on the quantile regression framework. It allows covariate adjustment, non-normally distributed traits, and continuous predictors (i.e., not limited to three genotype categories) (Miao et al., 2022). However, to integrate information from K quantiles (usually $K = 100$), QUAIL's computation time is much longer than that of its competitors (BF, DRM, DGLM, and KW). Implementing QUAIL for genome-wide vQTL analysis is computationally challenging.

This work evaluates the performance of the abovementioned parametric or non-parametric methods in detecting vQTLs through Monte Carlo simulations. I performed a head-to-head comparison to assess the false positive rates (FPRs), power, and computation time across several popular vQTL methods. Furthermore, I apply these approaches to four lipid traits in the Taiwan Biobank (TWB) data, including high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglyceride (TG), and total cholesterol (TCHO). After identifying vQTLs for these four traits, I perform direct GxE and GxG analyses to demonstrate enriched interaction effects among vQTLs.

2 Materials and methods

2.1 Covariate-adjusted residuals

Let Y_j be the continuous trait of the j th individual, $G_{1,j}$ and $G_{2,j}$ be two indicator variables categorizing three genotype

groups of a SNP, and X_j be the C -length covariate vector of individual j (assuming C covariates should be adjusted). $G_{1,j}$ and $G_{2,j}$ can be coded as (0, 0), (1, 0), and (0, 1) for 0, 1, and 2 copies of a SNP's specific allele. According to previous studies, the FPR of detecting vQTLs may be inflated in the presence of a large SNP's main effect (Ronnegard and Valdar, 2011). Therefore, I remove SNP's main effect and covariates' effects from Y_j to ensure that the identification of vQTLs will not be confounded by other factors. Specifically, I extract the residuals of regressing Y_j on $G_{1,j}$, $G_{2,j}$, and X_j . With this preliminary procedure, all parametric or non-parametric vQTL tests have the potential to adjust for covariates such as age, sex, and ancestry principal components (PCs).

2.2 Non-parametric vQTL methods

Non-parametric vQTL methods are developed without the assumption of normally distributed traits. Therefore, they are generally more robust to trait distributions (Miao et al., 2022). Like many non-parametric statistical tests, they may be less powerful than parametric vQTL methods when the traits indeed follow normal distributions (Ronnegard and Valdar, 2012). However, they can have more valid results even when the traits are not normally distributed. In the following, I introduce two non-parametric methods, including the Kruskal-Wallis test (KW) and the quantile integral linear model (QUAIL).

2.2.1 Kruskal-Wallis test (KW)

KW is a non-parametric statistical test used to assess whether there is a significant difference between the medians of two or more independent groups. Because the KW test is not originally designed to detect differences in variances, I have to convert the data into a measure of "dispersion" before using KW. Let $e_{(ij)}$ be the "covariate-adjusted residual" (obtained from Section 2.1) of individual j with genotype i , \bar{e}_i be the median of "covariate-adjusted residuals" within genotype group i . The deviation between individual j 's covariate-adjusted residual and his/her group median is $D_{(ij)} = |e_{(ij)} - \bar{e}_i|$ (a measure of dispersion), where $j = 1, \dots, N$ and usually $i = 1, 2, 3$ (subscript j is sufficient to distinguish all N individuals; subscript i is supplementary to indicate the genotype group). Suppose $r_{(ij)}$ is the rank of $D_{(ij)}$ counting from all individuals across three genotype groups. The KW test statistic (Kruskal and Wallis, 1952) is listed in Equation 1 as follows,

$$KW = (N - 1) \frac{\sum_{i=1}^M n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^M \sum_{j=1}^{n_i} (r_{(ij)} - \bar{r})^2} \quad (1)$$

where N is the total sample size, n_i is the number of individuals in genotype group i , M is the number of genotype groups (usually $M = 3$), $\bar{r}_i = \sum_{j=1}^{n_i} r_{(ij)} / n_i$ is the average of ranks in genotype group i , $\bar{r} = \sum_{i=1}^M \sum_{j=1}^{n_i} r_{(ij)} / N = \frac{1}{2} (N + 1)$ is the average of ranks among all individuals. The KW test statistic (1) follows the chi-square distribution with the degrees of freedom $(M - 1)$. All analyses in this work were conducted using R (version 4.3.1). The KW test was implemented with the R function "kruskal.test" in the R environment.

2.2.2 Quantile integral linear model (QUAIL)

Although I have adjusted the covariates' effects in Section 2.1, I still present the statistical model shown in QUAIL's paper as follows (Miao et al., 2022). Through this model, people may get a complete understanding of Miao et al.'s methodology. Let the conditional quantile function (Q) of the trait Y at a SNP be Equation 2 as follows

$$Q_Y(\tau | genotype = i, \mathbf{X}) = \mu_\tau + i\beta_\tau + \mathbf{X}\alpha_\tau \quad (2)$$

where τ is the quantile ranging from 0 to 0.5, $i = 1, 2, 3$, μ_τ is the intercept, β_τ is the coefficient of the quantile regression, \mathbf{X} is the $N \times C$ matrix for C covariates, and α_τ is the C -length vector for covariate effects. QUAIL measures $(\beta_{1-\tau} - \beta_\tau)$ given $\tau \in (0, 0.5)$, i.e., the difference between the regression coefficients of the $(1 - \tau)^{th}$ and τ^{th} quantiles. QUAIL tests the hypothesis $H_0: \beta_{QI} = 0$ vs $H_1: \beta_{QI} \neq 0$, where β_{QI} in Equation 3 is the quantile-integrated effect, i.e.,

$$\beta_{QI} = \int_0^{0.5} (\beta_{1-\tau} - \beta_\tau) d\tau \quad (3)$$

β_{QI} is the integral of $(\beta_{1-\tau} - \beta_\tau)$ from $\tau = 0$ to $\tau = 0.5$, which can be estimated by summing the areas of K rectangles with a length of $(\beta_{1-\tau_k} - \beta_{\tau_k})$ (where $k = 1, \dots, K$) and a width of $(0.5 - 0)/K = 1/2K$. That is, β_{QI} can be estimated through $\hat{\beta}_{QI} = \sum_{k=1}^K (\hat{\beta}_{1-\tau_k} - \hat{\beta}_{\tau_k})/2K$. To have a more accurate estimation for β_{QI} , K is required to be a large number. In the QUAIL R code, Miao et al. used $K = 100$ as the default setting (Miao et al., 2022). The QUAIL R code was downloaded from GitHub (<https://github.com/qlu-lab/QUAIL>).

2.3 Parametric vQTL methods

Parametric vQTL methods are usually developed with the assumption of normally distributed traits (Marderstein et al., 2021; Young et al., 2018). Therefore, their performances are generally more sensitive to trait distributions (Miao et al., 2022). However, if the trait really follows the normal distribution, parametric tests can be more powerful than their non-parametric counterparts. In the following, I describe three parametric vQTL methods, including the deviation regression model (DRM), Brown-Forsythe test (BF), and the double generalized linear model (DGLM).

2.3.1 Deviation regression model (DRM)

To search for SNPs that are associated with phenotypic variability, Marderstein et al. regressed $D_{(ij)}$ on the minor allele count at each SNP (i.e., 0, 1, or 2) (Marderstein et al., 2021). The regression model is $D_{(ij)} = \alpha + \beta \cdot (i - 1) + \varepsilon_j$, where the predictor $(i - 1)$ is the minor allele count (i.e., 0, 1, or 2) and the random error term ε_j follows a normal distribution with a mean of 0 and a variance of σ_ε^2 . Because $D_{(ij)}$ is the deviation between individuals' covariate-adjusted residuals and their group medians, this approach is called the deviation regression model (DRM). DRM tests whether the regression coefficient β is statistically significant, i.e., $H_0: \beta = 0$ vs $H_0: \beta \neq 0$. The DRM R code was downloaded from GitHub (<https://github.com/drewmard/DRM>).

2.3.2 Brown-Forsythe test (BF)

The BF test statistic is listed in Equation 4 as follows,

$$BF = \frac{\sum_{i=1}^M n_i (\bar{D}_i - \bar{D})^2 / (M - 1)}{\sum_{i=1}^M \sum_{j=1}^{n_i} (D_{(ij)} - \bar{D}_i)^2 / (N - M)} \quad (4)$$

where $\bar{D}_i = \sum_{j=1}^{n_i} D_{(ij)} / n_i$ is the average of deviations in genotype group i , $\bar{D} = \sum_{i=1}^M \sum_{j=1}^{n_i} D_{(ij)} / N$ is the average of deviations in all individuals. The BF test statistic (4) follows the F distribution with the numerator degrees of freedom $(M - 1)$, and the denominator degrees of freedom $(N - M)$ (Brown and Forsythe, 1974). The BF test is equivalent to the ANOVA F test for regressing D_{ij} on two indicator variables that separate the three genotype groups of an SNP. Therefore, BF treats genotypes as a categorical scale (by using two indicator variables), while DRM regards genotypes as a continuous scale (by coding genotypes as 0, 1, or 2).

To implement the BF test, I used the R function "leveneTest" while specifying "center = median" in the "car" package (version 3.1-3) (Fox and Weisberg, 2018). The Levene's and the BF tests are both used to test the homoscedasticity across different groups. The BF test is more robust to outliers than Levene's test because it uses deviations from the group median, while Levene's test uses deviations from the group mean.

2.3.3 Double generalized linear model (DGLM)

DGLM fits two generalized linear models (GLM) that model genetic effects on the mean and dispersion of the covariate-adjusted residuals (Zhang and Bell, 2024). It regresses the covariate-adjusted residual $e_{(ij)}$ on the minor allele count at each SNP (i.e., 0, 1, or 2) while allowing the dispersion of $e_{(ij)}$ varies with the minor allele count. The regression model is $e_{(ij)} = \alpha + \beta \cdot (i - 1) + \varepsilon_j$, where the random error term ε_j follows a normal distribution with a mean of 0 and a variance of σ_i^2 (allowing each genotype group has its variance). The DGLM tests whether the variance of $e_{(ij)}$ varies with the predictor (i.e., the minor allele count at each SNP). The DGLM method was implemented with the "dglm" R package (version: 1.8.6).

2.4 Data from the Taiwan Biobank (TWB)

As of February 2024, 147,836 individuals aged 30 to 70 years have been genotyped whole-genome. The TWB performed genotype imputation with the IMPUTE2 software (v2.3.1) (Delaneau et al., 2013; Howie et al., 2009). To improve the imputation accuracy, the TWB combined 1,445 TWB individuals with whole-genome sequence data and 504 East Asians (EAS) from the 1,000 Genomes Phase 3 v5 as the reference panel (Wei et al., 2021). After completing the imputation, the TWB excluded variants with missing rates >5%, minor allele frequencies (MAFs) < 0.01%, or imputation information scores < 0.3, where 0.3 was usually adopted as the acceptable threshold for imputation quality (Kosugi et al., 2023). After this quality control filtering, we had 9,814,944 autosomal variants for analysis.

2.5 Simulation studies

I performed simulations to evaluate the type I error rates and power of the five abovementioned methods. I here focused on the simulations of detecting SNP1-by-SNP2 interactions. Nonetheless, the results can be generalized to GxE identification. I randomly selected four SNPs as SNP1, including rs34625133 on chromosome (chr.) 1, rs7870809 on chr. 9, rs7982209 on chr. 13, and rs140100 on chr. 22. The MAFs of these four SNPs were 0.1, 0.2, 0.3, and 0.4, respectively. They were, in turn (one by one), treated as SNP1. I aimed to evaluate the power of detecting SNP1-by-SNP2 interactions under various MAFs of SNP1.

I used 20,000 common SNPs (MAFs ≥ 0.05) on chr. 1 as SNP2 and generated traits as Equation 5,

$$Y_j = \beta_1 \text{SNP}_{1,j} + \beta_2 \text{SNP}_{2,j} + \beta_{\text{INT}} \text{SNP}_{1,j} \times \text{SNP}_{2,j} + \varepsilon_j, j = 1, \dots, 30000 \text{ (or 147836)} \quad (5)$$

where $\text{SNP}_{1,j} = 0, 1$, or 2 represented individual j 's minor allele count at SNP1, and $\text{SNP}_{2,j} = 0, 1$, or 2 was his/her minor allele count at SNP2. I considered two levels of sample size: $N = 30,000$ or $147,836$, in which $147,836$ was the number of TWB individuals with genotyping data. The coefficients β_1 and β_2 are main effects of SNP1 and SNP2, respectively. I investigated two situations: (1) with SNP main effects ($\beta_1 = \beta_2 = 0.3$); and (2) without SNP main effects ($\beta_1 = \beta_2 = 0$). Given an individual's genotypes, $\beta_1 \text{SNP}_{1,j} + \beta_2 \text{SNP}_{2,j} + \beta_{\text{INT}} \text{SNP}_{1,j} \times \text{SNP}_{2,j}$ is the deterministic part and ε_j is the stochastic part of the trait. Following Miao et al. (Miao et al., 2022), I considered three settings for the random error term ε_j : (1) a standard normal distribution; (2) a t distribution with the degrees of freedom 3 (a kurtotic distribution); (3) a chi-square distribution with the degrees of freedom 6 (a skewed distribution). Because ε_j is the only stochastic part of the trait, the distribution of the trait is determined by the distribution of ε_j . A kurtotic distribution such as the t distribution has heavier tails than the normal distribution. This phenomenon mimics the fact that more extreme values exist than the normal distribution. The chi-square distribution is right-skewed like the pattern of many phenotypes, such as cholesterol (Tharu and Tsokos, 2017) and body weight (Kozłowski and Gawelczyk, 2002). To fix the phenotypic variance explained by SNPs across different distributions of ε_j , I followed Miao et al. (Miao et al., 2022) to perform the z -score transformation on ε_j .

By specifying $\beta_{\text{INT}} = 0$, I plotted the QQ plot to examine the p -values under the null hypothesis. On the other hand, $\beta_{\text{INT}} = 0.3$ was assumed when evaluating the power of the vQTLs methods. To sum up, 12 scenarios (two levels of sample size \times two situations of main effects \times three kinds of trait distributions) were simulated to evaluate the performance of the five vQTL methods. Each scenario was simulated 20,000 times. The percentage of the variance explained by the interaction effect was approximately 0.3%, 0.6%, 0.9%, and 1.2% when the MAF of SNP1 was 0.1, 0.2, 0.3, and 0.4, respectively. The coefficients in Equation (5) (i.e., β_1 , β_2 , and β_{INT}) were used to generate phenotype Y . Instead of testing any coefficient, in simulations, I tested whether phenotypic variance varied with genotype groups.

Although Equation (5) is originally designed to test for GxG interactions, the results can be generalized to GxE identification. Without loss of generality, I can replace SNP_2 in Equation (5) with

an environmental factor having three possible levels: 0, 1, and 2. Therefore, the simulation results can be generalized to GxE identification.

3 Results

3.1 Simulation studies

3.1.1 False positive rates

Figure 1 presents the QQ plots (the left column) and power under $\alpha = 5 \times 10^{-8}$ (columns 2–5) for five vQTL tests ($N = 30,000$; with SNP main effects). Supplementary Figure S1 demonstrates the results given $N = 30,000$, and SNP main effects do not exist. Supplementary Figure S2 (with SNP main effects) and S3 (without SNP main effects) show the results when $N = 147,836$. The rank-based inverse-normal transformation (INT) can convert data to a normal distribution. However, previous vQTL studies found that this INT transformation led to inflation in false positive rates (FPR) (Marderstein et al., 2021; Miao et al., 2022; Wang et al., 2019). The FPR was calculated by dividing the incorrectly classified negatives by the total negatives. To evaluate INT's performance in our simulation setting, we equipped it with DGLM.

3.1.1.1 Parametric tests

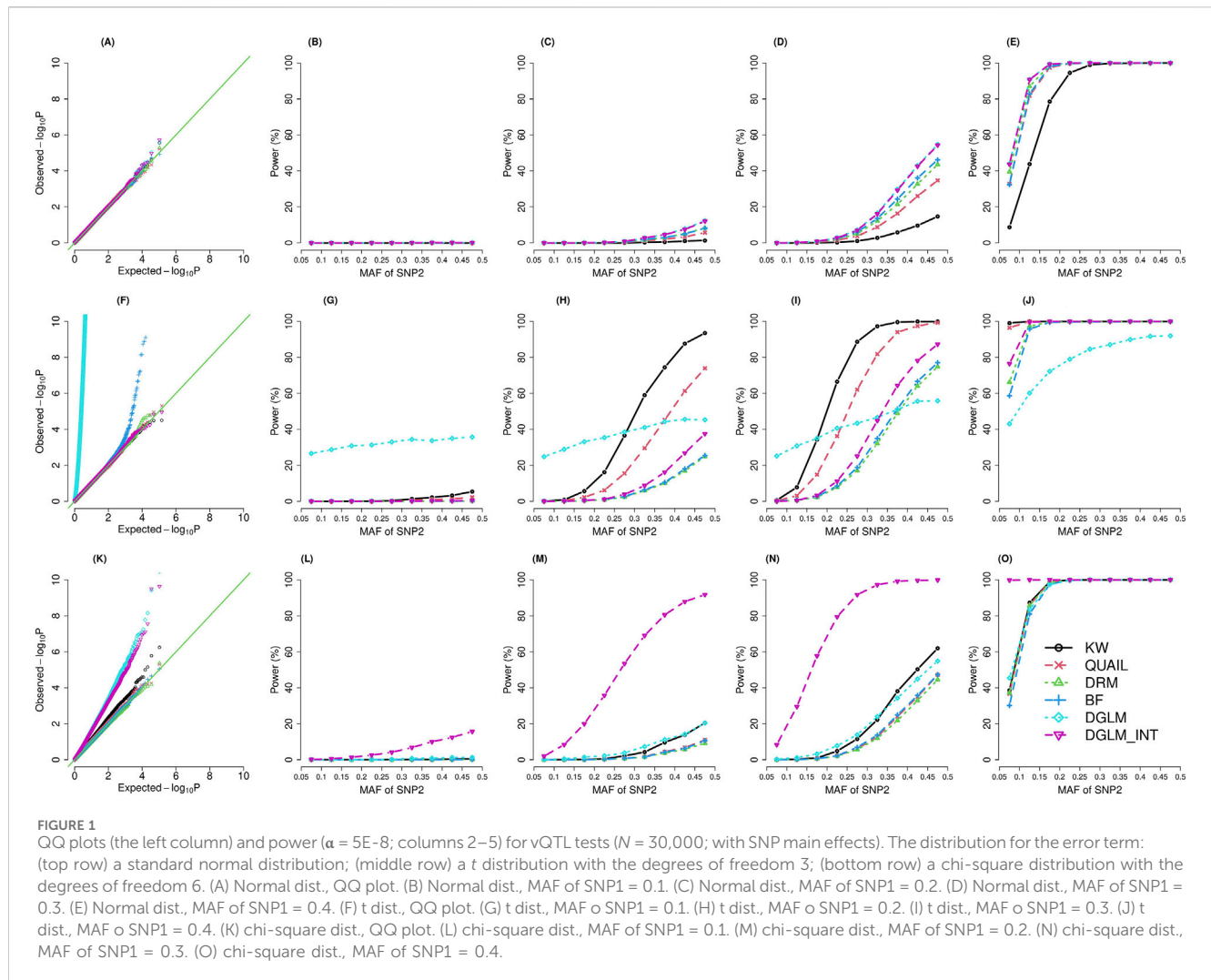
The results show that DGLM has inflated FPR when the phenotypes are not normally distributed (Figure 1; Supplementary Figures S1–S3F, K). Taking the INT transformation helps to adjust FPR for kurtotic phenotypes (Figure 1; Supplementary Figures S1–S3F) but not for skewed phenotypes (Figure 1; Supplementary Figures S1–S3K).

The BF test has inflated FPR given kurtotic phenotypes (Figure 1; Supplementary Figures S1–S3F), and the situation is getting worse for smaller sample sizes (Figure 1; Supplementary Figure S1F). Supplementary Figures S4–S15 demonstrate the QQ plots stratified by the MAF range. For the BF test, the inflation in FPR is especially severe for kurtotic phenotypes when $N = 30,000$ and $\text{MAF} < 0.2$ (Supplementary Figures S5, S8, (A) (B) (C)). Given $N = 30,000$ and a small MAF, a genotype group may only contain a few observations, and false positives may come with the issue of data sparsity (Lin, 2024).

DRM is similar to BF, except that DRM treats genotypes as a continuous scale (0, 1, or 2). DRM does not categorize the three genotypes into three groups. Therefore, the sparsity within genotype groups is not a critical problem for DRM, and the inflation in FPR is not critical for DRM compared with BF (Supplementary Figures S5, S8A–C).

3.1.1.2 Non-parametric tests

QUAIL preserves an appropriate FPR across three trait distributions. KW maintains a suitable FPR under the kurtotic distribution (Figure 1; Supplementary Figure S1–S3F) but has a slight inflation in FPR given the skewed distribution (Figure 1; Supplementary Figure S1–S3K). Under a kurtotic distribution, more extreme values can occur compared to the normal distribution. Nonetheless, by transforming the extreme values into ranks, KW can still preserve an appropriate FPR. By contrast, previous research has found that KW tends to have inflated FPR for heteroscedastic cases



(McDonald, 2014). Under a skewed distribution, unequal variances in ranks across the three genotype groups may still exist even if there is no GxG or GxE. KW's strategy to transform values into ranks cannot entirely address the heteroscedastic problem.

3.1.2 Statistical power

The statistical power was calculated by dividing the correctly classified positives by the total positives (each scenario was simulated 20,000 times). When the phenotype is normally distributed, DGLM and DGLM_INT were the most powerful methods (Figure 1; Supplementary Figure S1–S3C–E). Among the two non-parametric tests, QUAIL was superior to KW for normally distributed phenotypes (Figure 1; Supplementary Figure S1–S3C–E). When the phenotype is a kurtotic distribution, DGLM and BF had inflation in FPR (Figure 1; Supplementary Figure S1–S3F). KW became the most outstanding test among the four valid methods that preserve a suitable FPR (Figure 1; Supplementary Figure S1–S3F–J). When the phenotype is skewed, KW had a slight inflation in FPR, whereas DGLM and DGLM_INT had a severe inflation in FPR (Figure 1; Supplementary Figure S1–S3K). The other three valid methods (QUAIL, DRM, and BF) had similar power (Figure 1; Supplementary Figure S1–S3L–O).

By comparing the results of different simulation scenarios, I concluded three points: (1) When the MAF for SNP1 or SNP2 was larger, the percentage of the variance explained by the interaction effect increased, and the power of each method was enhanced. (2) When the sample size was enlarged from $N = 30,000$ to $N = 147,836$, the power performances of all methods were greatly improved. (3) Lastly, by comparing Figure 1 with Supplementary Figure S1, S2 with Supplementary Figure S3, the power of each method was higher in the presence of SNP main effects than in the absence of SNP main effects.

3.1.3 Computational time

Table 1 shows the computation time (in seconds) for each simulation replicate. I measured the execution time in R (version 4.3.1) on a Linux system running at 3.6 GHz and 32 GB of RAM. Although QUAIL is a valid test with appropriate FPR under three trait distributions, it takes much more computation time than other methods. The other non-parametric test, KW, requires only 1/70–1/90 of QUAIL's execution time. The computation time of the four parametric tests is also reasonable. The time of computation needed for these tests is ordered as $DRM < BF < KW < DGLM \approx DGLM_INT \lllll QUAIL$. Table 2 summarizes the performance

TABLE 1 Computation time (in seconds) for each simulation replicate The execution time was measured in R (version 4.3.1) on a Linux system running at 3.6 GHz and 32 GB of RAM.

Phenotype distribution	N	KW	QUAIL	DRM	BF	DGLM	DGLM_INT
Normal	30,000	0.076	5.372	0.018	0.020	0.151	0.162
T	30,000	0.092	8.252	0.024	0.028	0.273	0.214
Chi-square	30,000	0.075	5.485	0.018	0.021	0.163	0.168
Normal	147,836	0.402	33.464	0.096	0.105	0.824	0.891
T	147,836	0.399	33.950	0.094	0.103	1.090	0.875
Chi-square	147,836	0.411	35.913	0.098	0.106	0.884	0.929

TABLE 2 Summary of the simulation results.

	KW	QUAIL	DRM	BF	DGLM	DGLM_INT
False positive rate (FPR)	Slight inflation in FPR given skewed distributions	Valid	Inflation in FPR is not critical for DRM compared with BF.	Inflation in FPR is especially severe for kurtotic traits when N = 30,000 and MAF <0.2	Valid only when the trait is normally distributed	Severe inflation in FPR given skewed distributions
Statistical power	Optimal for kurtotic distributions	Never among the optimal choices	Similar to BF.	Similar to DRM.	Optimal for normally distributed traits	Optimal for normally distributed traits
Computational time (ranking from shortest to longest)	3	6 (longest)	1 (shortest)	2	4	4

of each method in the simulation on FPR, power, and computational time under different scenarios.

3.2 Real data analysis

3.2.1 30 vQTLs of four lipid traits

When conducting actual genome-wide data analysis, investigators had to exclude cryptic relatedness among individuals (Marees et al., 2018). The TWB investigated cryptic relatedness among participants with the KING (Kinship-based INference for GWAS) software (Manichaikul et al., 2010). I removed the person with higher missing genotype rates for each first- or second-degree relative pair. This step excluded 28,928 from the 147,836 TWB individuals, and 118,908 remained in the vQTL analysis.

I analyzed four lipid traits, including high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglyceride (TG), and total cholesterol (TCHO). Supplementary Figure S16 shows the histograms of the four lipid traits. All four traits are right-skewed with positive skewness values, especially TG (skewness = 2.16). I first removed SNP’s main effect and covariates’ effects from each phenotype. Specifically, I extracted the residuals of regressing each trait on $G_{1,j}$, $G_{2,j}$ (two indicator variables separating the three genotype groups of a SNP), and 17 covariates, including sex (male vs female), age (in years), BMI (in kg/m²), current smoking status (yes vs no), current drinking status (yes vs no), performing physical exercise (yes vs no), educational attainment (integer ranging from 1 to 7), and 10 ancestry PCs. Current smoking was

defined as “having smoked cigarettes for at least 6 months when joining the TWB”. Drinking indicated “having a weekly intake of more than 150 mL of alcoholic beverages for at least 6 months when joining the TWB”. Regular exercise meant “performing exercise lasting for 30 min thrice a week”. Educational attainment was coded as an integer ranging from 1 to 7: 1 (illiterate), 2 (no formal education but literate), 3 (primary school graduate), 4 (junior high school graduate), 5 (senior high school graduate), 6 (college graduate), and 7 (Master’s or higher degree).

Sex, age, and BMI significantly influence people’s lipid profiles (Beyene et al., 2020). Smoking impairs lipid metabolism by decreasing LDL receptor expression (Ma et al., 2020). Alcohol consumption disturbs lipid metabolism by increasing adipose tissue lipolysis and causing ectopic fat deposition in the liver (Steiner and Lang, 2017). Regular physical activity has been found to improve lipoprotein-lipid profiles (Lin, 2021). Moreover, lower educational attainment is usually linked to worse lipid profiles (Lara and Amigo, 2018). Because these seven covariates are associated with lipid traits, I adjusted them in all vQTL methods. Furthermore, the first 10 ancestry PCs are adjusted in genetic analyses to avoid population stratification (Uffelmann et al., 2021).

Based on the simulation results, I chose the KW test as the primary vQTL method. This non-parametric approach is robust to outliers by transforming trait values into ranks (Figure 1; Supplementary Figure S1–S3F). Moreover, although KW has a slight inflation in FPR for skewed trait distributions (Figure 1; Supplementary Figure S1–S3K), investigators can perform a follow-up regression analysis to check whether GxE and GxG

TABLE 3 Thirty vQTLs (KW p-value < 5E-8 and linkage disequilibrium measure r^2 < 0.01) of the four lipid traits.

Trait (number of vQTLs)	Chromosome	Base pair	vQTL	Gene
HDL (10)	9	104827463	rs4149307	<i>ABCA1</i>
	9	104902020	rs1883025	<i>ABCA1</i>
	15	58431476	rs1800588	<i>LIPC</i>
	16	56957451	rs183130	Near <i>CETP</i>
	16	67998971	rs113731015	<i>DPEP2</i>
	16	68256297	rs12446418	<i>PLA2G15</i>
	18	49627658	rs80208964	<i>LOC105372112</i>
	19	11236981	rs737338	<i>DOCK6</i>
	19	44912383	rs445925	<i>APOC1</i>
	20	45907572	rs148753678	<i>PLTP</i>
LDL (6)	2	21024193	rs57825321	<i>APOB</i>
	5	75325846	rs4704208 ^a	Near <i>HMGCR</i>
	5	75614147	rs11748027 ^b	<i>ANKDD1B</i>
	11	116786951	rs3741297	<i>ZPR1</i>
	19	11131631	rs2738464	<i>LDLR</i>
	19	44912921	rs483082 ^c	<i>APOC1</i>
TG (12)	1	62436136	rs631106	<i>USP1</i>
	1	62690372	rs1168114	<i>DOCK7</i>
	2	27518370	rs780094	<i>GCKR</i>
	2	27798099	rs3935148	<i>RBKS</i>
	7	73606007	rs3812316	<i>MLXIPL</i>
	8	19966163	rs1803924	<i>LPL</i>
	8	125465736	rs2001945	Near <i>TRIB1</i>
	11	61815236	rs174561	<i>FADS1</i>
	11	116792991	rs662799 ^d	<i>APOA5</i>
	11	117050674	rs1815786	<i>SIK3</i>
	17	67966122	rs10445361	<i>BPTF</i>
	19	44913484	rs438811	<i>APOC1</i>
TCHO (6)	1	109275536	rs3832016	<i>CELSR2</i>
	2	21029662	rs13306194	<i>APOB</i>
	5	75325846	rs4704208 ^a	Near <i>HMGCR</i>
	5	75614147	rs11748027 ^b	<i>ANKDD1B</i>
	11	116792991	rs662799 ^d	<i>APOA5</i>
	19	44912921	rs483082 ^c	<i>APOC1</i>

^ars4704208 (near *HMGCR*) is a vQTL of LDL and TCHO.

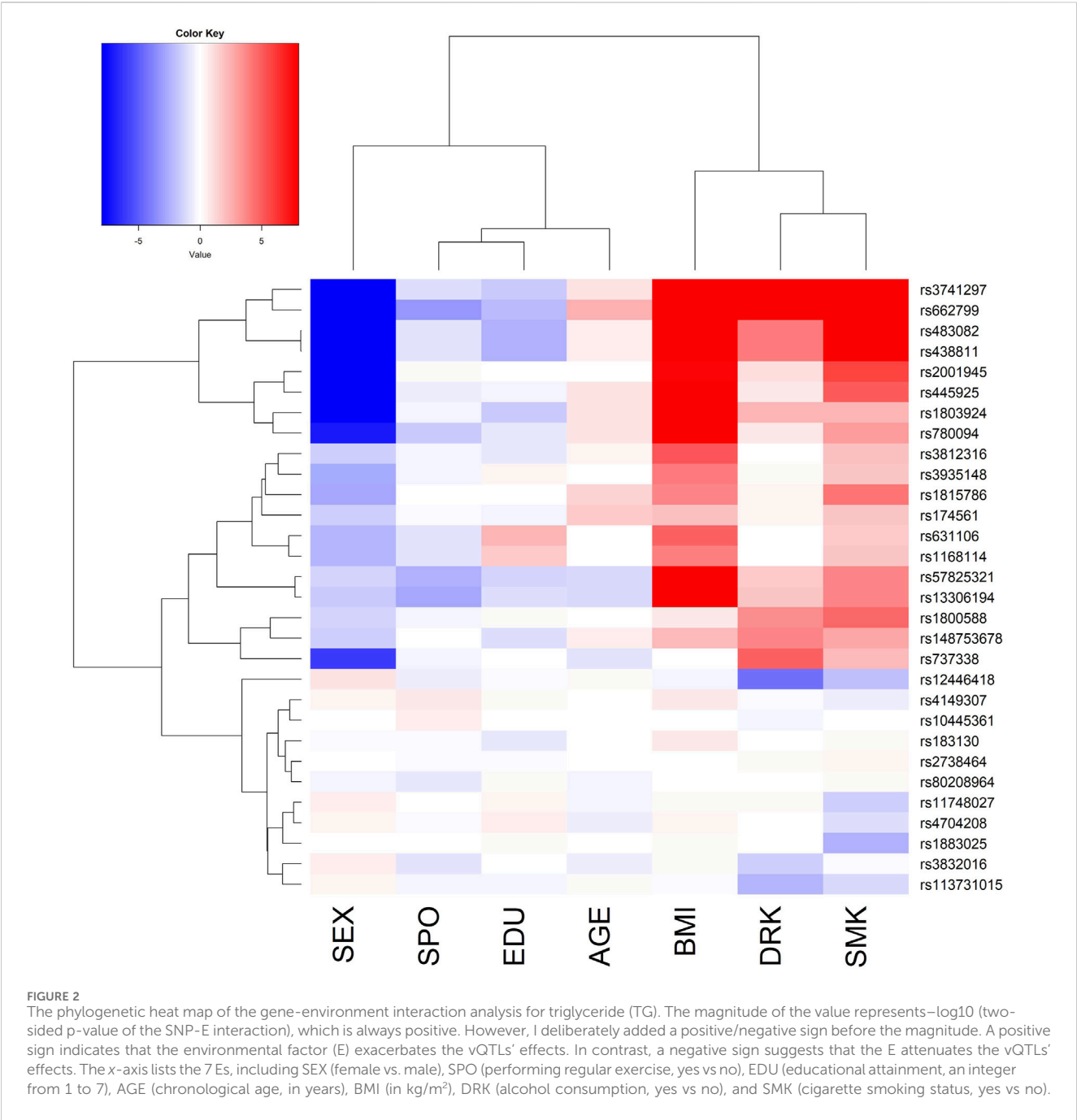
^brs11748027 (in *ANKDD1B*) is a vQTL of LDL and TCHO.

^crs483082 (in *APOC1*) is a vQTL of LDL and TCHO.

^drs662799 (in *APOA5*) is a vQTL of TG and TCHO.

exist (i.e., the so-called “direct GxE or GxG analysis”). Furthermore, KW needs only 1/70–1/90 of the other non-parametric competitor’s (QUAIL) execution time. Because SNPs with MAFs <0.05 are

difficult to replicate in GxG or GxE findings (Lin, 2024; Wang et al., 2019), I only analyzed 2,580,790 common SNPs with MAFs \geq 0.05.



With the PLINK clumping procedure (Purcell et al., 2007), I detected 10 independent HDL-vQTLs (KW p-value $< 5\text{E-}8$) with linkage disequilibrium (LD) measure $r^2 < 0.01$. Moreover, 6, 12, and 6 independent vQTLs were identified from LDL, TG, and TCHO, respectively. Four SNPs were found to be vQTLs of multiple lipid traits, including rs4704208 (near *HMGCR*), rs11748027 (in *ANKDD1B*), rs483082 (in *APOC1*), and rs662799 (in *APOA5*). Table 3 summarizes the 30 (=10 + 6+12+6-4) unique vQTLs of the four lipid traits.

3.2.2 Direct GxE analysis

From the regression (or variable selection) perspective, it is important to keep the hierarchical structure between main and

interaction effects (Zhou et al., 2021). Following previous vQTL research (Wang et al., 2019), after identifying vQTLs, I performed a “direct GxE analysis” by regressing the trait on the minor allele count of each vQTL (0, 1, or 2), seven environmental factors (Es), and the product (interaction) term between the minor allele count and one of the seven Es. To remove population stratification, I also adjusted the top 10 ancestry PCs in this regression. The seven Es are listed as the horizontal axis of Figure 2, including SEX (female vs male), SPO (performing regular exercise, yes vs no), EDU (educational attainment, an integer from 1 to 7), AGE (chronological age, in years), BMI (in kg/m^2), DRK (alcohol consumption, yes vs no), and SMK (cigarette smoking status, yes vs no). As Westerman et al. (Westerman et al., 2022) indicated, some



vQTLs were “pleiotropic” concerning phenotypic variability. Table 2 also shows that four loci are vQTLs shared by multiple lipid traits. Instead of performing the direct GxE and GxG analysis for respective vQTLs of each lipid trait (i.e., 10 vQTLs for HDL, 6 vQTLs for LDL, 12 vQTLs for TG, and 6 vQTLs for TCHO), I analyzed all 30 vQTLs to have a more comprehensive picture of the four lipid traits.

Because TG had more vQTLs compared with the other three traits, I present the phylogenetic heat map of the direct GxE analysis for TG in Figure 2. The x-axis of Figure 2 lists the 7 Es. The magnitude of the value in Figure 2 represents $-\log_{10}$ (two-sided p-value of the SNP-E interaction), which is always positive. However, I deliberately added a positive/negative sign before the

magnitude. A positive sign indicates that E exacerbates the vQTLs’ effects. In contrast, a negative sign suggests that the E attenuates the vQTLs’ effects. Females have more attenuated TG genetic effects than males (blue color in SEX column), whereas higher BMI, alcohol consumption, and cigarette smoking lead to more substantial TG genetic effects (red color in BMI, DRK, and SMK columns). Moreover, the phylogenetic heat maps for the remaining three traits are demonstrated in Supplementary Figures S17–S19.

3.2.3 Direct GxG analysis

I further performed a “direct GxG analysis” by regressing the trait on the minor allele counts (0, 1, or 2) of two SNPs, the seven Es mentioned above, and the product (interaction) term of the minor

allele counts of the two SNPs. Similarly, I adjusted the top 10 ancestry PCs in this regression. Figure 3 shows the phylogenetic heat map of the direct GxG analysis for TG. The maps for the other three lipid traits can be found in Supplementary Figures S20–S22.

Regarding TG, the interaction between *rs438811* (chr. 19, in *APOC1*) and *rs662799* (chr. 11, in *APOA5*) is the most significant among the $\binom{30}{2} = 435$ GxG tests constructed by the 30 vQTLs ($p = 1.3\text{E-}15$). The magnitude of the value in Figure 3 represents $-\log_{10}$ (two-sided p -value of the SNP-SNP interaction), which is always positive. The most significant SNP-SNP interaction is between *rs438811* (chr. 19, in *APOC1*) and *rs662799* (chr. 11, in *APOA5*), where $p = 1.3\text{E-}15$. SNP *rs483082* is highly correlated with *rs438811* ($r^2 = 0.99$), and *rs3741297* is in obvious linkage disequilibrium with *rs662799* ($r^2 = 0.19$). Therefore, despite four black cells in Figure 3 indicating four significant SNP-SNP interactions, I only highlighted *rs438811*-*rs662799* interaction. As analyzed by the direct GxG regression model, each T allele at *rs438811* is associated with 10.8 mg/dL ($p = 7.0\text{E-}117$), and each G allele at *rs662799* is associated with 26.4 mg/dL ($p \sim 0$, extremely low p -value). Possible values of the interaction (product) term between *rs438811* and *rs662799* are 0, 1 ($=1 \times 1$), 2 ($=1 \times 2$ or 2×1), and 4 ($=2 \times 2$). Each point of the interaction term is associated with 5.9 mg/dL ($p = 1.3\text{E-}15$). This result indicates a synergistic interaction between *rs438811* and *rs662799*: T allele at *rs438811* and G allele at *rs662799* working together exerts a more substantial impact on TG than they would have on their own.

4 Discussion

Searching for vQTLs greatly facilitates the mining of GxE and GxG. For a GWAS incorporating one million SNPs and seven Es, seven million GxE tests and $\binom{1000000}{2} = 499999500000$ GxG tests are required to investigate GxE and GxG. In the data analysis of this work, 30 vQTLs were identified from four lipid traits. I only needed to perform 210 ($= 30 \times 7$) GxE and 435 ($= \binom{30}{2}$) GxG tests for each trait. The number of tests has dramatically decreased, and the penalty for multiple tests can be attenuated with a preliminary vQTL search.

This work evaluates the performance of three parametric (DRM, BF, and DGLM) and two non-parametric (KW and QUAIL) methods in detecting vQTLs. Although DGLM is the most powerful method for normally distributed traits (Figure 1; Supplementary Figure S1–S3C–E), it has severe inflation in FPR for traits following other distributions (Figure 1; Supplementary Figure S1–S3F, K). The random error term of DGLM is assumed to follow the normal distribution. This is why it fails to control for FPR when the trait is non-normally distributed. Therefore, DGLM should not be adopted for vQTL detection when the trait distribution is unclear. Taking INT transformation helps adjust FPR for kurtotic traits (Figure 1; Supplementary Figure S1–S3F) but not for skewed traits (Figure 1; Supplementary Figure S1–S3K). Transforming trait values into ranks like INT and KW can address the outlier problem in kurtotic distributions. However, these two rank-based methods, especially DGLM_INT, cannot fully solve the

heteroscedastic problem under skewed traits (Figure 1; Supplementary Figure S1–S3K).

BF has inflated FPR under kurtotic traits when $\text{MAF} < 0.2$ (Supplementary Figures S5, S8A–C). The BF test is equivalent to the ANOVA F test for regressing D_{ij} on the two indicator variables differentiating the three genotype groups of an SNP. Therefore, BF is not a valid test given kurtotic traits and unbalanced sample sizes across three genotype groups (i.e., only a few observations in a particular genotype group, which is a common situation under small MAF). This unbalanced issue can be alleviated given a larger total sample size (N). As we can see, the problem of inflated FPR in BF is not so severe when N increases to 147,836 (Supplementary Figures S11, S14A–C).

DRM has a power performance similar to that of BF. These two tests are based on a similar strategy. DRM regresses $D_{(ij)}$ (the deviation between individual j 's covariate-adjusted residual and his/her group median) on the minor allele count at each SNP (i.e., 0, 1, or 2), whereas BF regresses $D_{(ij)}$ on the two indicator variables of genotypes. The only difference between the two methods is that DRM regards genotypes as a continuous scale (coding as 0, 1, and 2), whereas BF treats genotypes as a categorical scale. It is unsurprising that DRM and BF perform similarly in power. However, DRM controls FPR more appropriately than BF, because unbalanced sample size is a more critical issue when separating genotypes into three groups.

Two non-parametric tests, QUAIL and KW, were evaluated in this work. Although QUAIL maintains an appropriate FPR under normally or non-normally distributed traits (column 1 of Figure 1; Supplementary Figure S1–S3), its statistical power is never among the optimal choices under any situation (columns 2–5 of Figure 1; Supplementary Figure S1–S3). Besides, the power of QUAIL can be compromised by quantile crossing, which is a well-known issue when estimating multiple quantiles simultaneously (Bondell et al., 2010). However, the QUAIL R code does not evaluate whether an analysis suffers from quantile crossing. Moreover, QUAIL requires much longer computation time than the other competitors. Therefore, I would not recommend using QUAIL to perform a genome-wide vQTL search. On the other hand, KW is the most powerful vQTL test for kurtotic phenotypes (Figure 1; Supplementary Figure S1–S3G–J). When the traits are skewed, KW presents a slight inflation in FPR (Figure 1; Supplementary Figure S1–S3K). However, the subsequent direct GxE or GxG analysis can further assess the significance of GxE and GxG. Therefore, after considering the computation time, I chose KW to analyze the four lipid traits in the TWB data.

With the comprehensive simulations in this work, DRM and KW are the most recommended parametric and non-parametric vQTL tests, respectively. Unlike DRM (Marderstein et al., 2021), KW has no specific paper to introduce its implementation in detecting vQTLs. Therefore, as a demonstration, I here used KW to analyze the four TWB lipid traits. Other investigators may also choose DRM to perform genome-wide vQTL searches because of its adequate control of FPR and shorter computational time (Table 1). If one applies two or more methods to the same data set, he/she may explore the superiority of each method by comparing the replication rates of different techniques. That is, one can first separate the data set into a discovery set and a replication set, then calculate the replication rates of different methods.

Same with previous vQTL works (Lin, 2024; Wang et al., 2019), I only analyzed common SNPs with $\text{MAFs} \geq 0.05$. The reason is that GxE is the joint distribution between a SNP and E, and GxG is the

joint distribution between two SNPs. Interaction studies are difficult to replicate if an SNP has a small MAF (Lin, 2024). Figures 2, 3 show that GxE and GxG are enriched among vQTLs of the four lipid traits. With more vQTLs identified from TG, it is not surprising that more GxE and GxG can be explored from this trait (GxE: comparing Figure 2 with Supplementary Figures S17–S19; GxG: comparing Figure 3 with Supplementary Figures S20–S22).

For TG, *APOA5 rs662799* has been found to interact with cigarette smoking and alcohol consumption, based on individuals from Korea (Park and Kang, 2020). These interactions were also associated with metabolic syndrome (MetS) by analyzing subjects from Jilin Province of China (Wu et al., 2016). Because TG is one critical component of MetS, both studies (Park and Kang, 2020; Wu et al., 2016) were in line with our analysis results (*APOA5 rs662799* is on the second row of Figure 2).

DRM and KW are the most recommended parametric and non-parametric vQTL methods. QUAIL appropriately preserves the FPR under normally or non-normally distributed traits. However, its power is never among the optimal choices, and its computational time is much longer than the other competitors. BF suffers from severe inflation in FPR when SNP's MAF < 0.2. We may alleviate this inflation in FPR by increasing the sample size to, say, $N \sim 150,000$. DRM and BF are based on a similar strategy. DRM regresses the deviation between individuals' covariate-adjusted residuals and their group medians on the minor allele count of an SNP, whereas BF regresses the deviation between individuals' covariate-adjusted residuals and their group medians on two indicator variables categorizing three genotypes. Therefore, DRM and BF have similar power. However, BF suffers from more inflation in FPR than DRM does, mainly when a small sample size is observed in one genotype. Although DGLM is the most powerful method given normally distributed traits, it is not valid (i.e., producing large FPR) for non-normally distributed traits. Adopting the rank-based INT transformation can address the outlier problem and adjust the FPR for kurtotic traits. However, INT cannot fully solve the heteroscedastic issue of skewed traits.

Recently, a robust Bayesian mixed model rooted in the one-way ANOVA has been developed (Fan et al., 2025). By specifying a likelihood based on a heavy-tailed distribution, this method can improve robustness in GxE detection. With Fan et al.'s concept (Fan et al., 2025), the non-robust parametric tests can build a robust mixed-effect model.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.twbiobank.org.tw/new_web/, TWBR10810-07.

Ethics statement

The studies involving humans were approved by National Taiwan University Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. No potentially identifiable images or data are presented in this study.

Author contributions

W-YL: Formal Analysis, Methodology, Validation, Project administration, Data curation, Supervision, Writing – original draft, Conceptualization, Software, Funding acquisition, Investigation, Visualization, Resources, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the National Science and Technology Council of Taiwan (grant number 112-2628-B-002-024-MY3 to W.-Y.L.).

Acknowledgments

The authors thank the reviewers for their constructive comments and all the Taiwan Biobank investigators, staff, and participants for their contributions. This research has been conducted using the Taiwan Biobank Resource under application number TWBR10810-07.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1617504/full#supplementary-material>

References

- Beyene, H. B., Olshansky, G., Aa, T. S., Giles, C., Huynh, K., Cinel, M., et al. (2020). High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: evidence from two large population cohort studies. *PLoS Biol. Sep.* 18, e3000870. doi:10.1371/journal.pbio.3000870
- Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* 97, 825–838. doi:10.1093/biomet/asq048
- Breusch, T. S., and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287–1294. doi:10.2307/1911963
- Brown, M. B., and Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics* 16, 129–132. doi:10.1080/00401706.1974.10489158
- Delaneau, O., Zagury, J. F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6. doi:10.1038/nmeth.2307
- Fan, K., Jiang, Y., Ma, S. G., Wang, W. Q., and Wu, C. (2025). Robust sparse Bayesian regression for longitudinal gene-environment interactions. *J. Roy. Stat. Soc. C, qiaf027*. doi:10.1093/jrsssc/qiaf027
- Fox, J., and Weisberg, S. (2018). *An R companion to applied regression*. Thousand Oaks, United States: Sage Publications. Third Edition.
- Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *Plos Genet. Jun* 5, e1000529. doi:10.1371/journal.pgen.1000529
- Kosugi, S., Kamatani, Y., Harada, K., Tomizuka, K., Momozawa, Y., Morisaki, T., et al. (2023). Detection of trait-associated structural variations using short-read sequencing. *Cell. Genom* 14 (3), 100328. doi:10.1016/j.xgen.2023.100328
- Kozłowski, J., and Gawełczyk, A. T. (2002). Why are species' body size distributions usually skewed to the right?. *Funct. Ecol.* 16, 419–432. doi:10.1046/j.1365-2435.2002.00646.x
- Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47, 583–621. doi:10.2307/2280779
- Lara, M., and Amigo, H. (2018). Association between education and blood lipid levels as income increases over a decade: a cohort study. *BMC Public Health* 18, 256. doi:10.1186/s12889-018-5185-3
- Levene, H. (1960). Robust tests for equality of variances. In: *Contributions to probability and statistics; essays in honor of harold hotelling*. Redwood City, United States: Stanford University Press. p. 278–292.
- Lin, W. Y. (2021). A large-scale observational study linking various kinds of physical exercise to lipoprotein-lipid profile. *J. Int. Soc. Sport Nutr.* 18, 35. doi:10.1186/s12970-021-00436-2
- Lin, W. Y. (2024). Detecting gene-environment interactions from multiple continuous traits. *Bioinformatics* 40, btae419. doi:10.1093/bioinformatics/btae419
- Ma, B., Chen, Y., Wang, X., Zhang, R., Niu, S., Ni, L., et al. (2020). Cigarette smoke exposure impairs lipid metabolism by decreasing low-density lipoprotein receptor expression in hepatocytes. *Lipids Health Dis.* 19, 88. doi:10.1186/s12944-020-01276-w
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi:10.1093/bioinformatics/btq559
- Marderstein, A. R., Davenport, E. R., Kulm, S., Van Hout, C. V., Elemento, O., and Clark, A. G. (2021). Leveraging phenotypic variability to identify genetic interactions in human phenotypes. *Am. J. Hum. Genet.* 108, 49–67. doi:10.1016/j.ajhg.2020.11.016
- Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., et al. (2018). A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27, e1608. doi:10.1002/mpr.1608
- McDonald, J. H. (2014). *Handbook of biological statistics*. Baltimore, Maryland: Sparky House Publishing. 3rd Edition.
- Miao, J., Lin, Y., Wu, Y., Zheng, B., Schmitz, L. L., Fletcher, J. M., et al. (2022). A quantile integral linear model to quantify genetic effects on phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2212959119. doi:10.1073/pnas.2212959119
- Pare, G., Cook, N. R., Ridker, P. M., and Chasman, D. I. (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the women's genome health Study. *Plos Genet. Jun* 6, e1000981. doi:10.1371/journal.pgen.1000981
- Park, S., and Kang, S. (2020). Alcohol, carbohydrate, and calcium intakes and smoking interactions with APOA5 rs662799 and rs2266788 were associated with elevated plasma triglyceride concentrations in a cross-sectional study of Korean adults. *J. Acad. Nutr. Diet.* 120, 1318–1329. doi:10.1016/j.jand.2020.01.009
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Ronnegard, L., and Valdar, W. (2011). Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genet. Jun* 188, 435–447. doi:10.1534/genetics.111.127068
- Ronnegard, L., and Valdar, W. (2012). Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *Bmc Genet.* 13, 63. doi:10.1186/1471-2156-13-63
- Shi, G. (2022). Genome-wide variance quantitative trait locus analysis suggests small interaction effects in blood pressure traits. *Sci. Rep.* 12, 12649. doi:10.1038/s41598-022-16908-7
- Smyth, G. K. (1989). Generalized linear-models with varying dispersion. *J. Roy. Stat. Soc. B* 51, 47–60. doi:10.1111/j.2517-6161.1989.tb01747.x
- Steiner, J. L., and Lang, C. H. (2017). Alcohol, adipose tissue and lipid dysregulation. *Biomol. Mar.* 7, 16. doi:10.3390/biom7010016
- Tharu, B. P., and Tsokos, C. P. (2017). A statistical study of serum cholesterol level by gender and race. *J. Res. Health Sci.* Jul 25 (17), e00386.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Method Prime* 1, 59–21. doi:10.1038/s43586-021-00056-9
- Wang, H. W., Zhang, F. T., Zeng, J., Wu, Y., Kemper, K. E., Xue, A. L., et al. (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* 5, eaaw3538. doi:10.1126/sciadv.aaw3538
- Wei, C. Y., Yang, J. H., Yeh, E. C., Tsai, M. F., Kao, H. J., Lo, C. Z., et al. (2021). Genetic profiles of 103,106 individuals in the Taiwan biobank provide insights into the health and history of Han Chinese. *Npj Genom Med.* 6, 10. doi:10.1038/s41525-021-00178-9
- Westerman, K. E., Majarian, T. D., Giulianini, F., Jang, D. K., Miao, J., Florez, J. C., et al. (2022). Variance-quantitative trait loci enable systematic discovery of gene-environment interactions for cardiometabolic serum biomarkers. *Nat. Commun.* 13, 3993. doi:10.1038/s41467-022-31625-5
- Wu, Y. H., Yu, Y. Q., Zhao, T. C., Wang, S. B., Fu, Y. L., Qi, Y., et al. (2016). Interactions of environmental factors and APOA1-APOC3-APOA4-APOA5 gene cluster gene polymorphisms with metabolic syndrome. *Plos One* 11, e0147946. doi:10.1371/journal.pone.0147946
- Young, A. I., Wauthier, F. L., and Donnelly, P. (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* Nov. 50, 1608–1614. doi:10.1038/s41588-018-0225-6
- Zhang, X., and Bell, J. T. (2024). Detecting genetic effects on phenotype variability to capture gene-by-environment interactions: a systematic method comparison. *G3 (Bethesda)* 14, jkae022. doi:10.1093/g3journal/jkae022
- Zhou, F., Ren, J., Lu, X., Ma, S., and Wu, C. (2021). Gene-Environment interaction: a variable selection perspective. *Methods Mol. Biol.* 2212, 191–223. doi:10.1007/978-1-0716-0947-7_13